Methods

Hidden in the blow - a matrix to characterise cetaceans' respiratory microbiome: short-finned pilot whale as case study

Beatriz Santos[1,2*], Luís Afonso[1*], Filipe Alves[3,4], Ana Dinis[3,4], Rita Ferreira[3,4], Ana M. Correia[1,2], Raul Valente[1,2], Ágatha Gil[1,5,6], Luis Filipe C. Castro[1,2], Isabel Sousa-Pinto[1,2], Massimiliano Rosso[7,8], Cinzia Centelleghe[9], Sandro Mazzariol[9], Catarina Magalhães[1,2*], Maria Paola Tomasino[1*]

1 CIIMAR – Interdisciplinary Centre of Marine and Environmental Research, University of Porto, 4450-208, Matosinhos, Portugal
2 FCUP – Department of Biology, Faculty of Sciences, University of Porto, 4169-007, Porto, Portugal
3 MARE - Marine and Environmental Sciences Centre / ARNET - Aquatic Research Network, Agência Regional para o Desenvolvimento da Investigação Tecnologia e Inovação (ARDITI), 9020-105 Funchal, Portugal
4 Faculty of Life Sciences, University of Madeira, 9000-003 Funchal, Portugal
5 CITAB – Centre for the Research and Technology of Agro-Environmental and Biological Sciences, Department of Biology and Environment, University of Trás-os-Montes and Alto Douro, 5000-801, Vila Real, Portugal
6 IIM-CSIC – Institute of Marine Research of the Spanish National Research Council, 36208, Vigo, Pontevedra, Spain
7 CIMA Research Foundation - Centro Internazionale di Ricerca in Monitoraggio Ambientale, 17100, Savona, Italy
8 NBFC - National Biodiversity Future Center, 90133, Palermo, Italy
9 UNIPD - Department of Comparative Biomedicine and Food Science, University of Padua 35131, Padova, Italy

Corresponding authors: Catarina Magalhães (cmagalhaes@ciimar.up.pt); Maria Paola Tomasino (mtomasino@ciimar.up.pt)

Abstract

Cetaceans are key sentinel species of marine ecosystems and ocean health, being a strategic taxonomic group that evaluates the well-being of aquatic habitats and detects harmful environmental trends. Respiratory diseases are amongst the main causes of death in these animals, so identifying the microbiome community in their exhaled breath condensates (EBC), i.e. blow, has been proposed as a key biomarker for assessing respiratory health. Yet, to characterise microbiomes related to these animals' respiratory tract and use them as a proxy for health status, it is necessary to develop baseline data on the microorganisms associated with cetaceans. Here, the short-finned pilot whale (SFPW, *Globicephala macrorhynchus*) was used as a case study to validate the most suitable primer set to explore the prokaryotic diversity of the cetaceans' respiratory tract. DNA extracted from blow samples (n = 12) of animals off Madeira Island was sequenced to amplify both V3-V4 and V4-V5 hypervariable regions of the 16S rRNA gene, using the same sequencing platform (Illumina MiSeq). Independently of the primer set used, all blows shared Actinobacteria, Bacteroidetes, Firmicutes and Proteobacteria phyla in their composition. V3-V4 resulted in a higher diversity of taxa with relative abundance above 1%, whereas the V4-V5 primers captured a higher number of microbial Amplicon Sequence Variants, detecting the rare microbial biosphere with pathogen potential. Additionally, it captured the core microbiome more efficiently. Thus, this study provides a detailed characterisation of SFPW respiratory-associated microbial communities, strengthening the idea of sociality influencing microbiome composition in the respiratory tract. Moreover, it supports the use of blow as a relevant biomarker for the physiological state of the airways in free-ranging cetaceans.

Key words: blow sampling, exhale breath condensate, *Globicephala macrorhynchus*, health status, metabarcoding, respiratory disease

---

* These authors contributed equally to the work.

Introduction

As keystone species, cetaceans play vital ecological functions, considering their role as nutrient vectors, position in the food chains and use as bioindicators of environmental health (Ballance 2018). Currently, multiple global threats pose a serious concern to the conservation of these aquatic mammals (Evans 2018), affecting individual health and ultimately compromising population viability (Nicol et al. 2020). Addressing their impacts on cetacean populations is, therefore, crucial yet challenging, as many species are elusive and inhabit remote habitats with wide distributions (Parsons et al. 2015).

Respiratory tract infections make a substantial contribution amongst the primary causes of death observed in these marine mammals (Díaz-Delgado et al. 2018; Cuvertoret-Sanz et al. 2020). Therefore, the identification of the microbiome community existent in their exhaled breath condensates (EBC), i.e. blow, has been proposed as a key biomarker for assessing respiratory health and to provide information for conservation management (Acevedo-Whitehouse et al. 2010). Cetacean blow is a relatively unexplored subject (Mello and Oliveira 2016), though more recent studies study it as a biological matrix (Atkinson et al. 2021; O'Mahony et al. 2024). It is mostly composed of the host's DNA, hormones, respiratory microbes, a range of metabolites and compounds related to inflammation and the immune system (Hunt et al. 2013). Several studies have reported important differences in the microbial communities detected in the blow of cetaceans when compared to those occurring in the external environment, thus being considered specific to cetaceans' respiratory tract (Pirotta et al. 2017; Geoghegan et al. 2018; Vendl et al. 2021). Within this material, potential specific microbial pathogens have been identified (Acevedo-Whitehouse et al. 2010), highlighting the value of using the blow collection and characterisation as a method for monitoring respiratory microbial communities in cetaceans (Lima et al. 2012).

Previous research on the airway microbiota of cetaceans resorted to the analysis of blow samples, using different sampling methodologies, DNA extraction and amplification and targeting different hypervariable gene regions, some of these using a metabarcoding approach. Nevertheless, the majority of data accessible on cetacean-associated microbiome, namely pathogens, diseases and parasites, come from captive, stranded, sick or injured individuals, which cannot be considered representative of the free-ranging populations (Johnson et al. 2009; Acevedo-Whitehouse et al. 2010; Lima et al. 2012). Therefore, although some studies have provided insights from free-ranging baleen whales (Apprill et al. 2017), knowledge of the respiratory microbiome of free-ranging cetaceans is rather limited. Here, we focus on short-finned pilot whales (*Globicephala macrorhynchus*; SFPW), a thoroughly studied species due to its global distribution, abundance and propensity to mass strandings (Betty et al. 2023). Living in social groups with long-lasting relationships and exhibiting distinct levels of site fidelity to specific areas (Boran and Heimlich 2019; Aguilar de Soto and Alves 2023), SFPW represents an interesting case-study of how social complexity might shape microbiome composition.

To the best of our knowledge, there are no published studies on the SFPW respiratory microbiome. In addition, no previous studies on other cetacean species have compared the amplification targeting different hypervariable gene

regions using the same methodology on the same samples. In light of this, the present study addresses the respiratory microbiome of free-ranging cetaceans, using SFPW as a model species. The main goal is to provide a network of consistent microbiome core taxa of the SFPW blow, by comparing V3-V4/V4-V5 hypervariable regions of 16S rRNA gene. The V4−V5 regions are currently recommended to target marine microbes (including both bacteria and archaea) by the Earth Microbiome Project (Walters et al. 2015). On the other end, the primers 341F and 806R, used in this study to amplify the V3-V4 regions of the 16S rRNA bacterial and archaeal genes, were already applied in a previous blow study (Centelleghe et al. 2020). Other prior studies analysing the blow microbiota have amplified the regions V3-V4, but used different primers in relation to the present study (Bik et al. 2016; Robles-Malagamba et al. 2020). The expected outputs will serve as a baseline to set a working and optimised methodology for SFPW health monitoring, from sample collection to laboratory analysis.

Methods

Blow Sampling

Blow sampling was conducted in September and October 2018, during at-sea research monitoring campaigns in the southern waters of Madeira Island, Portugal, targeting SFPW (Fig. 1). In addition to blow collection, the natural behaviour (travelling, resting, feeding or socialising), age class (following Betty et al. (2023); Aguilar de Soto and Alves (2023)) and the number of individuals sampled in the group (Suppl. material 1: table S1) were recorded.

Sample collection was carried out using a PERFORMAgene™ PG-100 swab collection kit (DNA Genotek®). This kit was attached to an extendable 5-metre aluminium pole and used as a sampling device to collect the blow (Suppl. material 1: fig. S1). As described by the company (www.dnagenotek.com), the "PERFORMAgene is a simple all-in-one non-invasive swab kit for the collection, stabilisation and transportation of animal DNA samples", being very practical to handle. Additionally, with this kit, the samples are stable for a year under room temperature conditions, which allows for some safe flexibility in terms of temperature fluctuation during fieldwork and transportation processes. An 8-metre rigid inflatable boat with an outboard 150 horsepower engine was used to slowly approach the animals (under permits 508/2018 and 10661/2018 from the Instituto de Florestas e Conservação da Natureza IP-RAM) and when they surfaced to exhale, the blow collector device was positioned about 40 to 50 cm above the blowhole into the exhaled plume to collect the droplets in the blow. Simultaneously, the sampled animals were photographed to distinguish between sampled individuals and to compare with the (OOM/MARE-ARDITI) Madeira's photographic-identification (ID) catalogue of the species following Alves et al. (2019). The latter allowed confirming if they belong to the island-associated population of short-finned pilot whales in Madeira (i.e. regularly captured in different seasons throughout the years; following Alves et al. (2013, 2020)) or if they were transient (i.e. captured only once), with the categories detailed in the Suppl. material 1: table S1. Nevertheless, the residency pattern was attributed as a reference and no statistical comparison was made due to the low number of individuals/samples in each category.
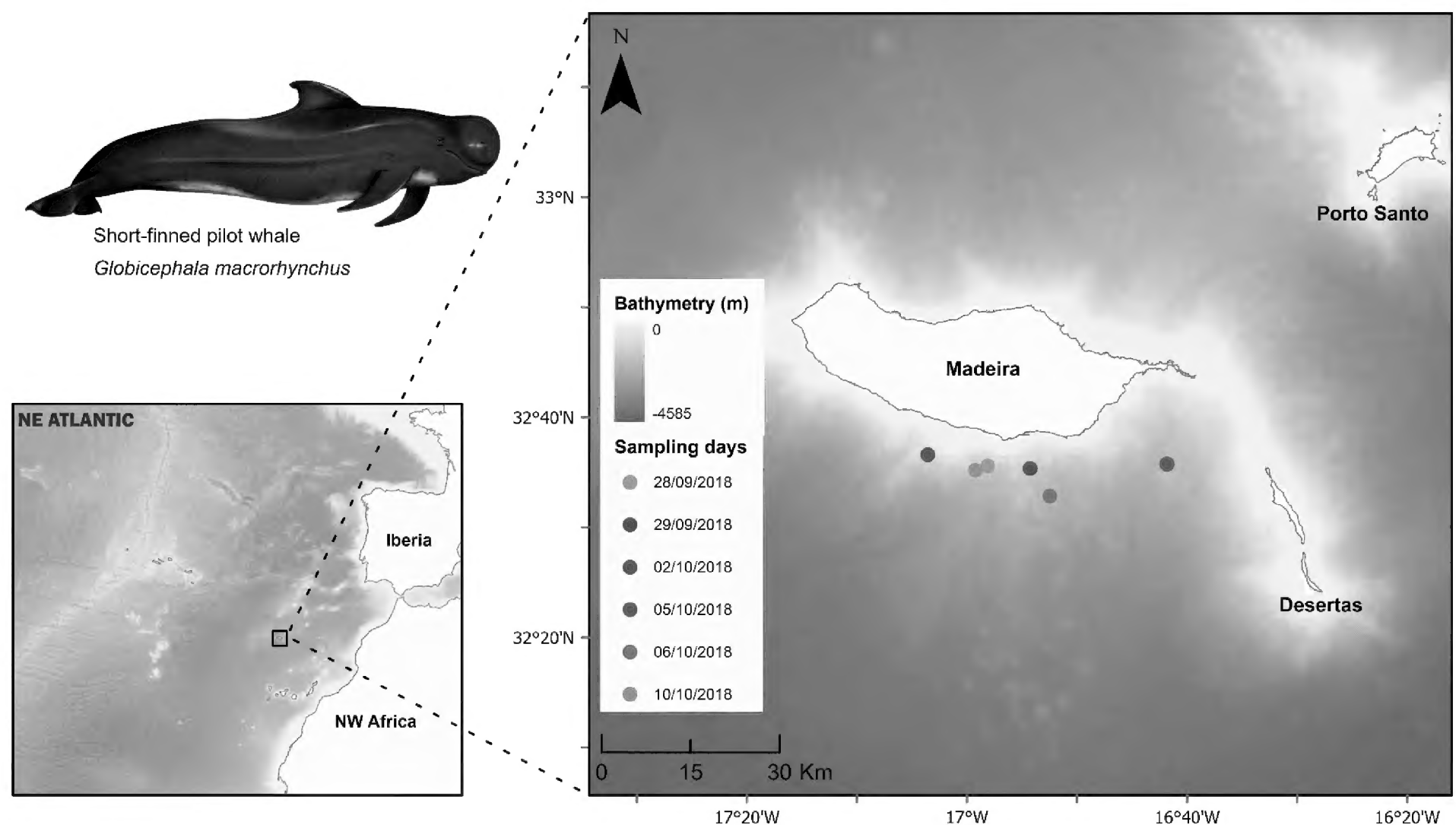
Figure 1. Blow sampling locations and dates in the Madeira Archipelago, with an illustration of the short-finned pilot whale (*Globicephala macrorhynchus*; © E. Berninsone / ARDITI).

The samples were kept refrigerated after collection, during transportation and until storage in the laboratory. Under lab conditions, samples were kept frozen until extraction. The maximum time between collection and extraction was approximately one year.

DNA extraction

DNA extraction from blow samples was performed using the QIAmp® DNA Mini Kit (QIAGEN), following the manufacturer's instructions. An additional sample concentration step in an Eppendorf Concentrator Plus™ was added to increase the concentration of the extracted DNA from blow samples. A Qubit™ 3 Fluorometer with a Qubit™ dsDNA High Sensitivity (HS) assay kit (Invitrogen™) was used for DNA quantification, after running negative (0 ng/µl dsDNA) and positive (10 ng/µl dsDNA) controls.

PCR amplification, library preparation and sequencing

Samples were prepared for the amplification of the 16S rRNA gene hypervariable V4-V5 region ($\approx$ 400 base pairs (bp)), using specific primers 515F-Y/926R (Parada et al. 2016). The V3-V4 region ($\approx$ 300 base pairs (bp)) of the 16S rRNA bacterial and archaeal gene was amplified using the 341F and 806R primers (Takahashi et al. 2014). Primer sequences can be found on Suppl. material 1: table S1.

Amplification and sequencing of both regions were performed through a dual-step PCR protocol followed by high-throughput sequencing. The PCR reactions included 2.5 µl of template DNA in a total volume of 25 µl. PCR conditions involved a 3 min denaturation step at 95 °C, followed by 35 cycles of 98 °C for

20 s, 60 °C for 30 s and 72 °C for 30 s and, finally, an extension stage at 72 °C for 5 min. A second PCR reaction was performed to add indexes and sequencing adapters to the target region, according to manufacturer's recommendations (Illumina 2013). Negative controls without templates were included in all PCR reactions. A detailed description of the protocol is reported in Ribeiro et al. (2018). Briefly, the KAPA HiFi HotStart PCR Kit was used in the first PCR reaction following manufacturer's suggestions. Indexes and sequencing adapters were added to the target amplicon in the second PCR reaction. The amplified products obtained were purified and normalised with a SequalPrep 96-well plate kit (ThermoFisher Scientific, USA). Pair-end sequencing was carried out in the Illumina MiSeq® sequencer with the V3 chemistry (Illumina, Inc., San Diego, CA, USA) at Genoinseq laboratories (Cantanhede, Portugal).

Bioinformatic analysis

Upstream analysis

The *"DADA2"* software package (v.1.20) (Callahan et al. 2016) on R studio (v.4.1.1) was used to process the FASTQ files obtained after Illumina MiSeq sequencing. Initially, the raw sequences were quality-filtered. Raw reads were truncated at 275 bp and 215 bp for forward and reverse sequences, respectively – the position where the final sequence numbers were higher. Afterwards, the trimmed sequences were dereplicated, denoised and merged. Sequences were then processed to obtain the Amplicon Sequence Variants (ASVs) table for taxonomic resolution. In the ASVs, sequences differ to the level of a single-nucleotide (Callahan et al. 2017). Additionally, chimeric sequences were identified and excluded. To determine the taxonomic classification of each ASV, the Naïve Bayes classifier was used against the SILVA database v.132 (Quast et al. 2013). SILVA is a curated taxonomic database that provides comprehensive, quality-checked and regularly updated datasets of aligned 16S ribosomal RNA sequences and is predominantly used for microbial molecular identification.

Downstream analysis

The ASV counts and taxonomy tables from the upstream analysis, together with the metadata table containing the sample information (Sample ID, sampling day and geographical coordinates and residency pattern), were used as input for the *"phyloseq"* R package (v.1.36) for downstream analysis (McMurdie and Holmes 2013). In summary, *"phyloseq"* analyses and graphically displays complex phylogenetic sequencing data already clustered into ASVs. The ASVs that were taxonomically unclassified at phylum rank or were not assigned to bacterial or archaeal lineages and the undesirable lineages, such as "Chloroplast", "Eukaryota" and "Mitochondria", were excluded from further analysis.

The distribution and diversity of the prokaryotic community across the different used primer sets were investigated. Alpha diversity was analysed for different subsets of samples by calculating four different indexes, also using *"phyloseq"* package: Observed and Chao 1 for the estimation of unique ASVs abundance and Shannon and Inverse Simpson as species diversity measures. Differences in alpha diversity were tested using the Wilcoxon signed rank test,

considering a level of significance of 0.05. Moreover, the beta diversity was analysed using the "*vegan*" package and a Non-metric MultiDimensional Scaling (NMDS) plot, based on the Bray-Curtis dissimilarity ("*phyloseq*" package). This measure is a statistical index used to quantify the compositional dissimilarity between two different sites, depending on the two communities' counts of shared and non-shared taxa (Bray and Curtis 1957). A PERMANOVA statistical test was implemented to test if the set of primers used had a significant effect on the prokaryotic communities in the different blow samples. Therefore, two hypotheses were generated: H_0: The use of different primer sets does not influence the distribution of prokaryotic communities; H_1: The use of different primer sets influences the distribution of prokaryotic communities. The level of significance was set to 0.05.

The microbial communities´ taxonomic composition was evaluated by creating taxonomy bar plots representing the relative abundance of each identified taxa (number of reads per taxa per total read count in each sample). The distribution of prokaryotes taxa across the samples was analysed at four taxonomic levels: phylum, class, family and genus. Taxonomic distribution plots were generated using several R-packages, namely, "*phyloseq*" (v.1.36) (McMurdie and Holmes 2013), "*ggplot2*" (v.3.3.6) (Wickham 2009), "*tidyverse*" (v.1.32) (Wickham 2009) and "*scales*" (v.1.20) (Wickham et al. 2019).

Core microbiome

To identify the common set of microbial taxa that originated for each of the datasets, the core microbiome was calculated using the "*phyloseq*" package "*microbiome*" (Lahti and Shetty 2017). The core microbiome was determined using the "*core_members*" function, considering a detection threshold of 0.001 and a prevalence of 50% in all blow samples. The comparative analysis of the primers used to identify this core microbiota was then carried out using a NMDS plot, based on Bray-Curtis dissimilarity.

Results

Blow sampling characterisation

A total of 12 blow samples were collected from six sampling events (Fig. 1), of which nine were from individual animals and three from a pool of animals (Table 1, Suppl. material 1: table S2). The photographic-ID comparison showed that the nine individual samples corresponded to seven individuals, i.e. two samples were obtained from the same animals in two separate sampling events (Table 1, Suppl. material 1: table S2).

Sequencing output

After the sequencing process, a total of 685,692 reads (with a mean of 57,141 ± 7,920.8 reads per sample) was obtained for the V3-V4 dataset; and 862,763 for the V4-V5 (with a mean of 71,896.9 ± 12,990.5 reads per sample). After the quality filter steps, the V3-V4 dataset showed a lower decrease in the number of sequences throughout the workflow compared to V4-V5, with 60.4 ± 5.7%

Table 1. Summarised metadata of blow samples collected and analysed. * Samples gma_01a and gma_01b correspond to the same individual, as samples gma_07a and gma_07b; ** Samples gma_P1, gma_P2, gma_P3 were collected from a pool of individuals within the same group.

| Sample ID	Sampling date	Class age	Type of sample
gma_01a *	28/9/2018	Adult	Isolated individual
gma_01b *	28/9/2018	Adult	Isolated individual
gma_02	29/9/2018	Adult	Isolated individual
gma_03	29/9/2018	Adult	Isolated individual
gma_04	2/10/2018	Adult	Isolated individual
gma_05	2/10/2018	Adult	Isolated individual
gma_06	5/10/2018	Adult	Isolated individual
gma_07a *	6/10/2018	Adult	Isolated individual
gma_07b *	6/10/2018	Adult	Isolated individual
gma_P1 **	6/10/2018	Adult	Pool
gma_P2 **	6/10/2018	Adult	Pool
gma_P3 **	10/10/2018	Adult	Pool

and 48.5 ± 7.7% of sequences retained per sample, respectively. Therefore, the non-target sequences were eliminated. Consequently, the final output was 415,340 sequences in the V3-V4 dataset that were assigned to 2,764 ASVs (with a mean of 34,611.7 ± 6,228.9 reads per sample) and 417,733 sequences in the V4-V5 dataset that were assigned to 3,665 ASVs (with a mean of 34,811.1 ± 7,554.7 reads per sample).

Alpha and beta diversity

Alpha diversity metrics varied with the set of primers used (Fig. 2). The Observed ASVs and Chao1 had lower minimum/maximum and mean values in the V3-V4 dataset relative to values of the V4-V5 dataset. On the other hand, alpha indices (Shannon and Inverted Simpson) values were higher for the V3-V4, although with an evident high standard deviation. For the V4-V5 dataset, despite lower values for these aforementioned diversity indices, there was less variation within samples (lower standard deviations). In the V4-V5 dataset, sample Gma_03 was an outlier, presenting the highest values in Observed and Chao1 indexes. The Shannon and Inverse Simpson indexes demonstrated more ASVs diversity in the V3-V4 region. The Wilcoxon Signed-Rank Test - used to compare the blow community richness for the Observed, Chao1 and Inverse Simpson measures - was significantly different between the two primers sets (see Suppl. material 1: table S3), with ca. 32% more bacterial ASVs in the V4-V5. Statistically significant differences were not found for the Shannon index.

The NMDS plot for beta diversity of microbial communities depicted a separation of the samples, based on the used primer set, with four distinct main clusters, two for each primer set (Fig. 3). One of the clusters of the V3-V4 dataset was composed of samples Gma_01a, 02, 03, 04 and 06 and the other cluster with samples Gma_01b, 07a, 07b, P1, P2 and P3. Gma_05 was an outlier for this dataset in this analysis. In the V4-V5 dataset, one of the clusters was
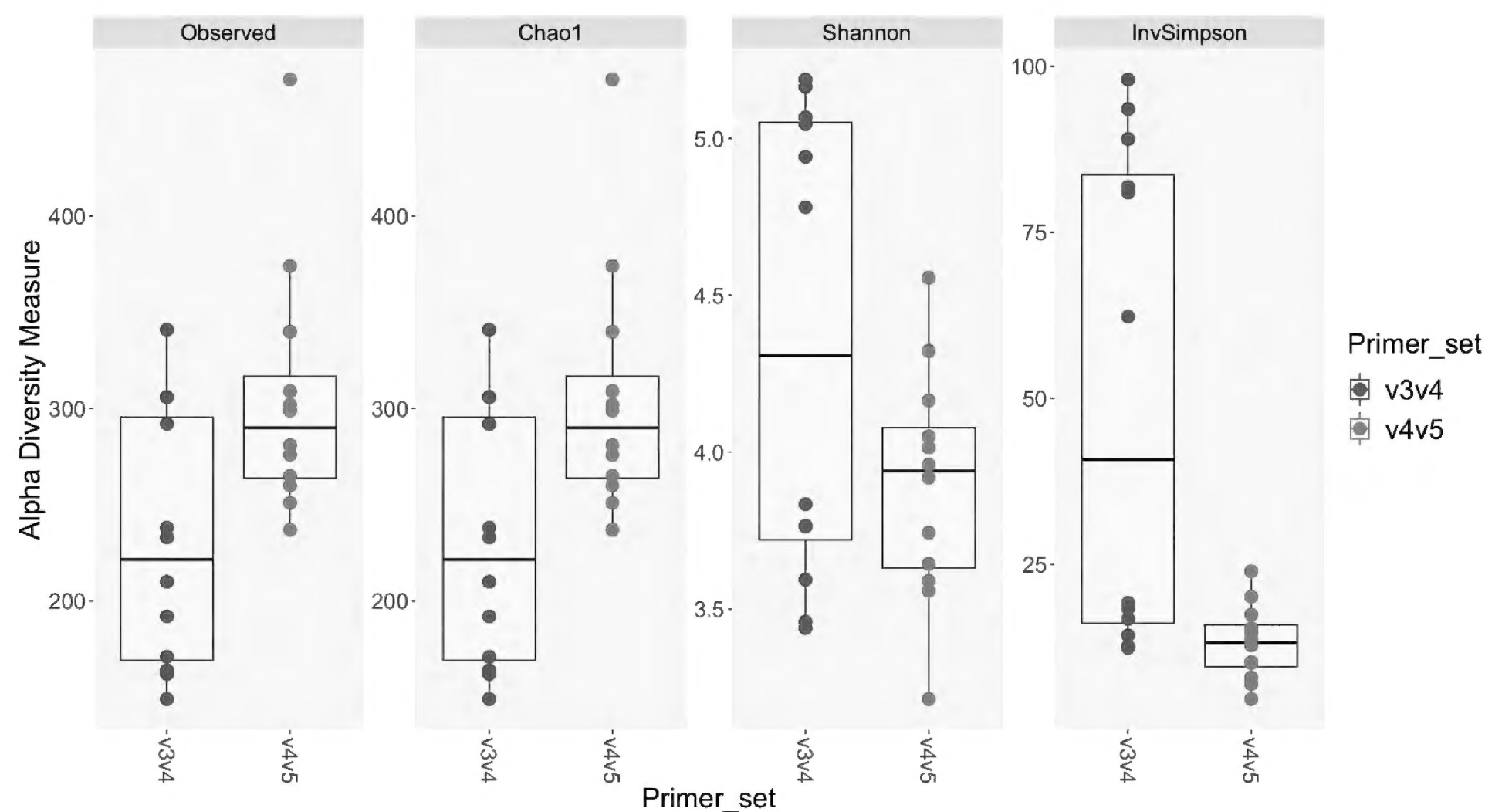
Figure 2. Box plot diagrams representing the median, first quartile and third quartile of observed richness, Chao1, Shannon and Inverse Simpson alpha diversity indices of the blow microbiome. Different primer sets were represented by different colours (blue for V3-V4 and orange for V4-V5).
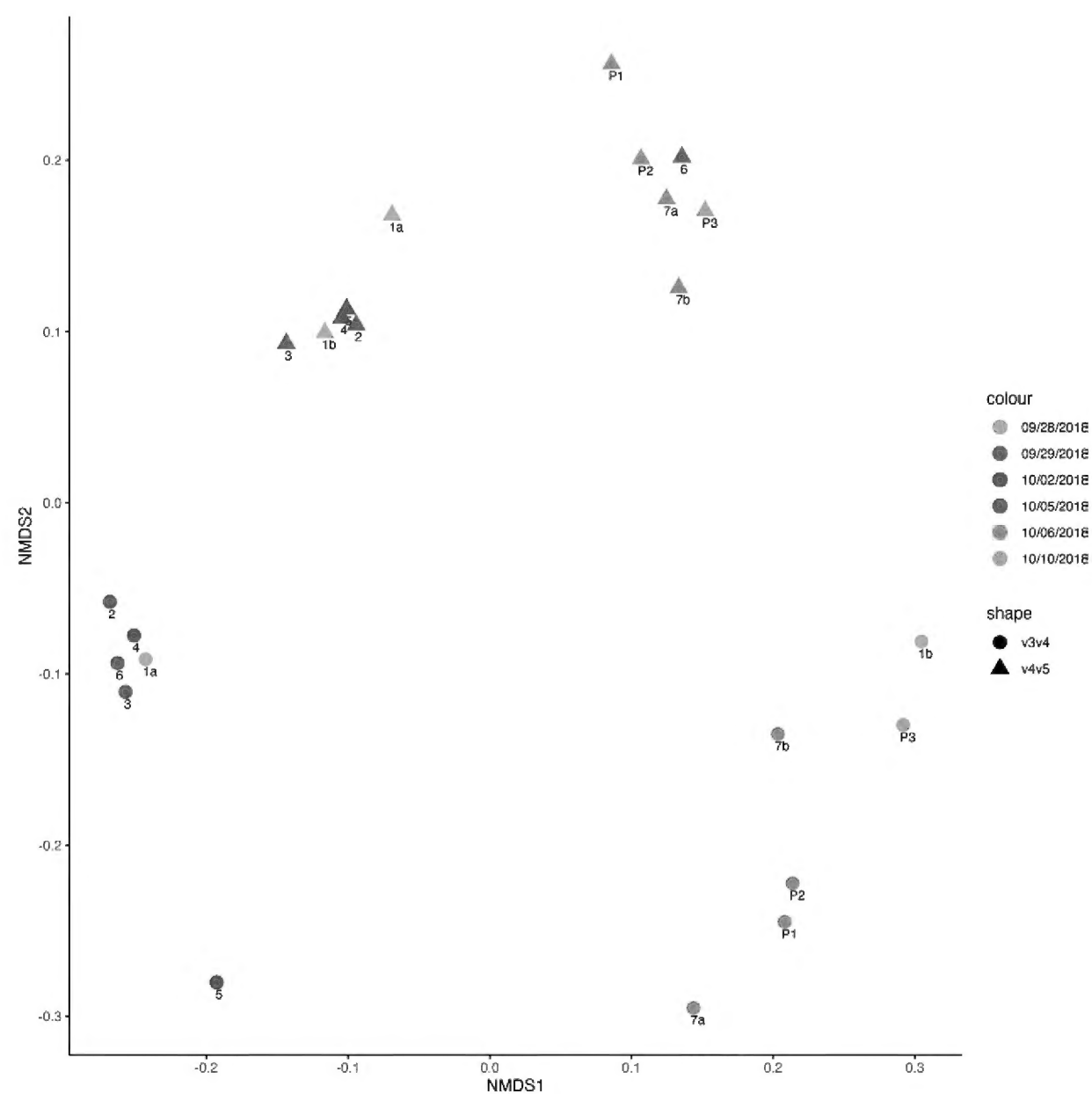
Figure 3. Non-metric multidimensional scaling (NMDS) of the microbiota found in short-finned pilot whale blow samples, in merged V3-V4 and V4-V5 datasets, based on Bray-Curtis dissimilarity. The primer sets used are represented by different shapes (circles for V3-V4 and triangles for V4-V5) and the sampling days by different colours.

composed of Gma_01a, 01b, 02, 03, 04 and 05 samples and the other one of Gma_06, 07a, 07b, P1, P2 and P3. Blow samples from individuals travelling together clustered in the same group for the V4-V5 dataset, which was not the case for the V3-V4 dataset. The most similar values were obtained for the samples of the individuals Gma_04 and 05 (sampled on the same day and in the same group of animals), within the V4-V5 dataset.

Regarding the PERMANOVA test results, the primer set influenced the prokaryotic communities' distribution (p = 0.006). Specifically, approximately 25% (R^2 = 0.249) of the prokaryotic community distribution variation was explained by using different primers (Suppl. material 1: table S4).

Blow taxonomic characterisation

The composition of the blow prokaryotic communities was investigated on the two datasets (V3-V4 and V4-V5) to analyse the differences and commonalities between using the two primer sets. Considering the richness, 1,890 bacterial and 38 archaeal taxa were recorded in the V3-V4 dataset, whereas V4-V5 recorded 2,327 bacterial and 30 archaeal taxa. Taxonomic analysis at the phylum level (Fig. 4A) revealed a core of four most abundant phyla identified by the two datasets: Actinobacteria, Bacteroidetes, Firmicutes and Proteobacteria. In the V3-V4 dataset, the prokaryotic community was distributed across 15 different phyla, seven of them had more than 1% of the total proportion of all the ASVs sequences of this dataset. This dataset was dominated by sequences of the phylum Proteobacteria (47%). On the other hand, V4-V5 was distributed within 21 phyla, with only five phyla with more than 1% of the total proportion of the V4-V5 dataset. This dataset was dominated by Proteobacteria (37%) and Actinobacteria (34%). Cyanobacteria was present in more samples of the V4-V5 dataset.

At the Class level (Suppl. material 1: fig. S1), the V3-V4 dataset included 37 different classes (11 of them were highly abundant, with > 1% of the sequence reads) and it was dominated by Gammaproteobacteria (31%), followed by the same proportion of Actinobacteria and Alphaproteobacteria (15%). The V4-V5 dataset was dominated by the Actinobacteria class (34%), with a low abundance of the majority of the sequence reads (representing < 1%) (30 out of 36). The V4-V5 dataset was mainly composed of six classes (Actinobacteria, Gammaproteobacteria, Alphaproteobacteria, Bacilli, Bacteroidia and Oxyphoyobacteria), whereas the V3-V4 dataset had in its composition the same taxa plus Gracilibacteria, Campylobacteria, Saccharimonadia and Clostridia. Within the classes Actinobacteria, Alphaproteobacteria, Bacteroidia, Clostridia, Gammaproteobacteria, Gracilibacteria, Parcubacteria and Saccharimonadia, there were many differences between datasets in the number of observed ASVs. These classes belong to the top 10 of both datasets, except Clostridia (position 11 of the V3-V4 dataset), Gracilibacteria (position 13 of the V4-V5 dataset) and Saccharimonadia (position 27 of the V4-V5 dataset).

Regarding the family level (Suppl. material 1: fig. S2), the prokaryotic community was distributed amongst 108 families and 24 families comprised more than 1% of the total proportion. This dataset was dominated by sequences of the family Pseudomonadaceae (18%), whereas V4-V5 was distributed by 153 families and only 17 families with more than 1% of the total proportion. Although the V4-V5 dataset presented more families, most had a relative abundance of <
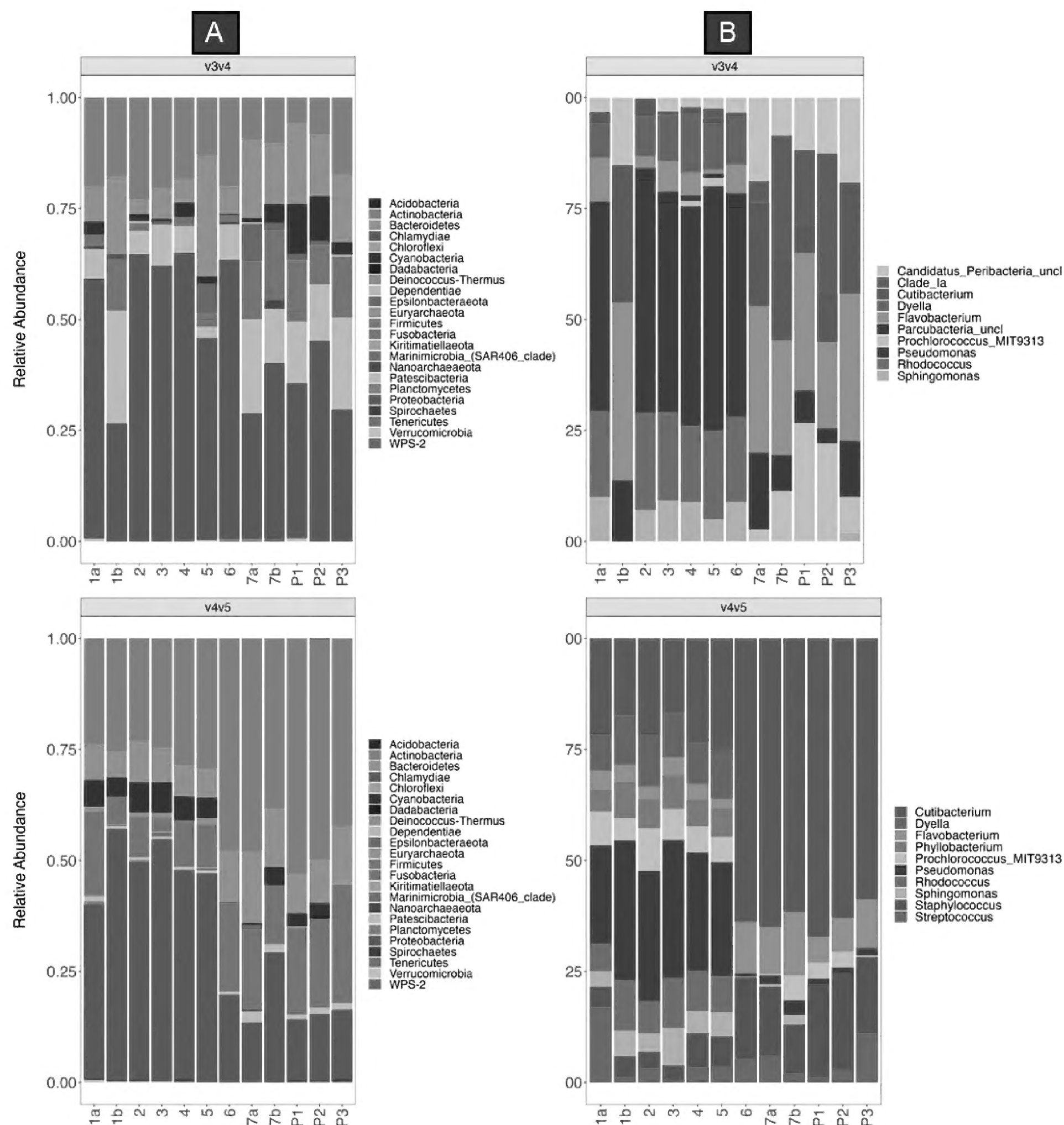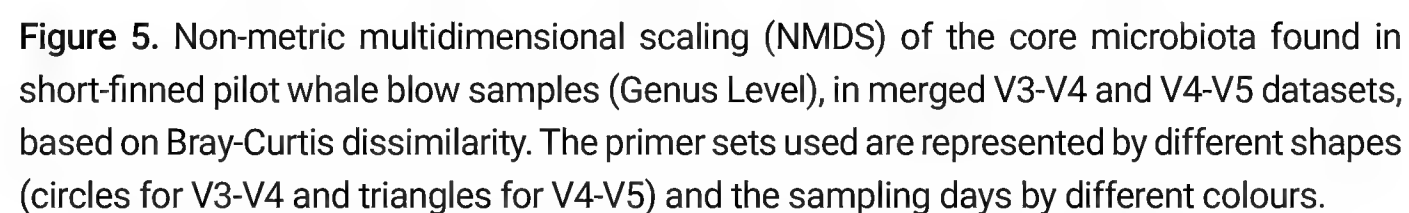
Figure 4. Relative abundance of prokaryotic phyla (**A**) and top 10 genera (**B**) identified across the SFPW blow samples, by V3-V4 (on the top) and V4-V5 (on the bottom) datasets.

1%. Propionibacteriaceae (25%) was the predominant family of this dataset. A comparison between the different datasets regarding the sequence proportion of the major taxonomic families identified can be found in the supplementary material (Suppl. material 1: fig. S3).

Concerning the genus level (Fig. 4B), the V3-V4 dataset detected 24 genera (with more than 1% of the total proportion of the sequences) associated with 128 different ASVs. This dataset was dominated by sequences of the genus *Pseudomonas* (18%), with only three ASVs associated with this genus. On the other hand, the V4-V5 dataset had only 14 genera (with more than 1%

of the total proportion) associated with 79 different ASVs and dominated by *Cutibacterium* (25%). The top 10 genera included some common taxa to both datasets: *Cutibacterium*, *Dyella*, *Flavobacterium*, *Prochlorococcus_MIT9313*, *Pseudomonas*, *Rhodococcus* and *Sphingomonas*. In the V3−V4 dataset, the ASVs were merged into 104 different genera and 53 lineages affiliated with higher taxonomic ranks. In the V4−V5 dataset, the ASVs were merged into 184 different genera and 68 lineages affiliated with higher taxonomic ranks. Overall, at this level, 110 (36.8% of the total) lineages were observed in both datasets. Considering all the taxa, in the V3−V4 dataset, 47 (15.7% of the total) lineages were absent from the V4−V5 dataset. Moreover, in the V4−V5 dataset, 142 lineages (47.5% of the total) were absent from the V3−V4 dataset.

At the family and genus levels, it is possible to observe two clusters of samples in terms of composition: Gma_01a, 02, 03, 04, 05, 06 and Gma_01b, 07a, 07b, P1, P2, P3 in V3-V4; Gma_01 to 05 and Gma_06 to P3 in V4-V5 (Suppl. material 1: fig. S2, Fig. 4B, respectively).

Core microbiome

The NMDS for the core microbiome analysis demonstrated an obvious separation of the samples based on the primer set used, making it possible to differentiate three distinct main clusters: two originated from the V3-V4 dataset and a unique cluster for the V4-V5 dataset (Fig. 5).

Figure 5. Non-metric multidimensional scaling (NMDS) of the core microbiota found in short-finned pilot whale blow samples (Genus Level), in merged V3-V4 and V4-V5 datasets, based on Bray-Curtis dissimilarity. The primer sets used are represented by different shapes (circles for V3-V4 and triangles for V4-V5) and the sampling days by different colours.

Discussion

Sampling methodology

The main goal of non-invasive sampling techniques for cetaceans is to avoid disturbing, injuring or negatively influencing the sampled individual during sample collection (Robinson and Nuuttila 2020). Apparently, none of the SFPW sampled here was negatively impacted by the sampling procedure, although, for future work, we suggest designing methodologies parallel to sampling to monitor the animals' behavioural responses, as tested by O'Mahony et al. (2024). This is based on no evident changes in the animals' swimming patterns or anomalous behaviour during sampling, suggesting that the blow-sampling method causes minimal distress to the individuals.

The sampling kit used (PERFORMAgene) has been employed in previous studies and applied to different animals, such as livestock (Neary et al. 2014) and pets (Sacco et al. 2017; Colpitts et al. 2022). This kit is suitable for genotyping, sequencing, parent-child appraisal and biobanking (Foley et al. 2011). Until now, no studies inliterature have reported using the PERFORMAgene kit to sample the blow of free-ranging cetaceans. In the present study, this kit was tested for the first time to collect blow samples from SFPW. DNA was extracted from the swab of the PERFORMAgene kit and, although low concentrations were obtained (Suppl. material 1: table S2), the method allowed for the employment of metabarcoding analysis.

Comparison of primer sets

In this study, the prokaryotic community harboured in the respiratory tract of SFPW was described for the first time and a comprehensive comparison of the performance of two different 16S primer sets' (V3-V4 and V4-V5) was conducted. Both hypervariable 16S regions are recommended in literature for assessing marine microbial diversity (Klindworth et al. 2013; Walters et al. 2015).

Using different hypervariable regions, as well as different types of samples, storage, methods of DNA extraction and 16S databases, can influence the obtained data and the interpretation of the results. In McNichol et al. (2021), that compared commonly-used primers, with > 300 million rRNA gene sequences retrieved from marine metagenomes around the world, the best-performing primers, when comparing predicted median coverage to Bacteria and Archaea of 16S rRNA, were 515Y/926R (regions V4-V5) and 515Y/806RB (region V4). The V4–V5 regions are currently recommended to target marine microbes (including bacteria and archaea) (Walters et al. 2015). On the other hand, prior studies analysing the blow microbiota of cetaceans have amplified the regions V3-V4 (Bik et al. 2016; Robles-Malagamba et al. 2020). Specifically, Centelleghe et al. (2020) used the primers 341F and 806R (as in this study) to amplify the V3-V4 regions of the 16S rRNA bacterial and archaeal genes. One of the aims of the present research was to assess the advantages and limitations in the amplification of both V3-V4 and V4-V5 regions regarding prokaryotic diversity and richness without the influence of technical biases (same experimental and bioinformatic processes on the same samples). Moreover, with this work, we also intended to select an optimal PCR primer set (or different applications of each primer set), that can be applied to the study of the microbiome of cetacean

blow samples. The results denote that the different hypervariable regions tested provide different degrees of resolution in taxonomic identification, resulting in different estimates of microbial community composition.

Alpha diversity measures, captured by each primer set, differed significantly. Our results showed that the V4-V5 dataset captured more abundance in unique ASVs (higher values in the Observed ASVs and Chao1 measures) and, subsequently, more identified taxa. On the other hand, V3-V4 resulted in higher values for the applied alpha diversity indexes (higher values in Shannon and Inverse Simpson measures). This is probably explained by the fact that these used indexes only consider unique ASVs in relative abundance above 1%. Beta diversity revealed that samples were grouped according to the primer sets used. Despite separating the samples in different clusters in relation to the primer set used, the V4-V5 dataset appeared to better represent the distribution of prokaryotic communities. In this dataset, individuals that were travelling together when they were sampled (namely Gma_2/Gma_3, Gma_4/Gma_5 and Gma_7/Gma_P1/Gma_P2) appeared within the same cluster and had more approximate values in the NMDS when compared to the V3-V4 dataset. There is evidence that sociality affects microbes in the respiratory tract (Vendl et al. 2020), which seems to be the case for the SFPW sampled in the present work. Some types of behaviour, such as surfacing and breathing near each other or feeding cooperatively, could facilitate the transfer of microbes and the spread of pathogens between individuals (Bogomolni et al. 2008; Apprill et al. 2017). Such types of behaviour are common in highly social and matrilineal species, such as the SFPW (Olson 2009; Boran and Heimlich 2019), as demonstrated in the target population (Alves et al. 2013; Esteban et al. 2022) to which the sampled animals belong. These processes have been recognised as exclusive and important aspects of social living, providing animal health benefits and acting as a driving force in the evolution of social behaviour (Lombardo 2008). Nevertheless, this hypothesis needs further evidence, since the V3-V4 dataset does not corroborate these results, possibly not reflecting the true blow prokaryotic community composition. Vendl et al. (2020) targeted solely the V4 region to study the microbiome in the blow of different whale species and observed a species-specific clustering in the microbiome beta-diversity, detecting a positive correlation between sociality and microbial diversity. Analysing the microbial composition in the breath of a representative number of individuals with different levels of site fidelity to the study area - which could not be statistically assessed in the present study due to a low number of individuals from each category - would be an interesting future line of enquiry, particularly as this would further shed light on the effect of sociality and habitat use on breath microbiota. Additionally, this type of comparative study could be paired with other predictors of health, such as photogrammetry-sourced measures of body condition.

The analysis of the blow core microbiome could provide useful features for the health monitoring of cetaceans worldwide. All samples from both datasets shared a main core microbiota in their blow, composed of Actinobacteria, Bacteroidetes, Firmicutes and Proteobacteria phyla. Nevertheless, the dominant ASVs were not the same between the results obtained from the amplification of V3-V4 and V4-V5 regions. Lima et al. (2012) described a temporal analysis of the blow in captive dolphins, suggesting that microbial community

composition in healthy animals is quite stable and that individual dolphins harbour consistently unique microbial communities. Several studies provided preliminary evidence that cetaceans host a core group of bacteria associated with the respiratory system (Johnson et al. 2009; Lima et al. 2012; Bik et al. 2016; Apprill et al. 2017). Apprill et al. (2017) detected, on average, 25 sequence clusters in 100% of humpback whales, accounting for 36% of the microbiota, being one of the most extensive core microbiotas found in any mammal. This large core microbiome was shared across individual whales from populations separated geographically into two ocean basins. Vendl et al. (2019) found low microbiota richness and a small core microbiome in humpback whales off the coast of Australia. Those sampled animals had been migrating for four months, which is the fasting period; whereas the animals sampled in the study by Apprill et al. (2017) were at their foraging sites and in their early stages of migration. Vendl et al. (2019) hypothesised that the lack of a core microbiome might be related to the animals' physiological state at the sampling time.

In this study, at a phylum level, the most dominant taxa recovered from both datasets were Proteobacteria. This aligns with previous studies of other species of cetaceans, either of baleen whales (Apprill et al. 2017; Pirotta et al. 2017; Vendl et al. 2019; Atkinson et al. 2021) and of dolphins w(Nelson et al. 2019; Centelleghe et al. 2020; Atkinson 2021) where this taxon is highly abundant. Lima et al. (2012) showed that the aforementioned common phyla, here present in both datasets (Proteobacteria, Actinobacteria, Bacteroidetes and Firmicutes), plus the Fusobacteria and unclassified bacteria, represented 98% of the total community composition in bottlenose dolphins. In the present work, Fusobacteria was also detected, but in lower abundance (< 1% of the total sequence reads) in both datasets, contrasting with the results from Nelson et al. (2019) where this taxon accounted for 7,2% of the total community. The presence of Cyanobacteria, a seawater phylum, in blow samples, is not surprising, having been detected in previous studies of the cetacean blow microbiota (Pirotta et al. 2017). Control water samples were not taken in this study, but we consider this to be extremely important for future work, in order to analyse the extent to which the detected microbial composition of the blow samples is influenced by seawater at the time of sampling. Geoghegan et al. (2018) and Centelleghe et al. (2020) have previously reported the presence of prokaryotes in blow samples that are also common to the microbiome found in surface layers of seawater. In both studies, the authors demonstrate a lower abundance of common prokaryotic taxa between the microbial composition of the blow and seawater, also showing a clear distinction between them. Epsilonbacteraeota, a phylum with a marked presence in the V3-V4 dataset, has already been identified as part of the core microbiota of bottlenose dolphins' respiratory system (Lima et al. 2012). In the current study, we found the phyla Euryarchaeota, Dadabacteria, Kiritimatiellaeota, Acidobacteria and Deinococcus-Thermus exclusively in the V4-V5 region, which were not previously reported in studies investigating the blow microbiota of various cetacean species. Similarly, Dependentiae in the V3-V4 dataset has not been documented in prior research. Contrarily, Planctomycetes, Chloroflexi (Bik et al. 2016) and Spirochaetes (Lima et al. 2012; Bik et al. 2016; Vendl et al. 2021), here only found in the V4-V5 dataset, were described in other studies concerning blow microbiota.

Regarding the genera level, *Cutibacterium* (dominant in the V4-V5 dataset) is a typical dominant microbial community in the nasal microbiota (Kumpitsch et al. 2019). The genus *Prochlorococcus*, detected in both datasets, is one of the most abundant bacteria in the ocean (Zehr et al. 2017) and it is probably a seawater-associated bacterium within the SFPW blow. Four of the most abundant genera detected in both datasets (*Pseudomonas*, *Flavobacterium*, *Rhodococcus* and *Sphingomonas*) are potential pathogenic genera within the blow microbiome, commonly related to different diseases, such as various pulmonary infections (Higgins 2000; Venn-Watson et al. 2008; Venn-Watson et al. 2012; Apprill et al. 2017). This is not necessarily indicative of disease since not all organisms belonging to these taxa are pathogenic. The V4-V5 dataset detected more diversity of the less abundant taxa in the microbiome composition of the blows, therefore highlighting the value of this primer set to detect the "rare biosphere" in the blow. This can be especially relevant, considering the low-represented taxa that can potentially have a pathogenic role in causing respiratory diseases. Different potentially pathogenic genera were less abundant in the blow samples analysed in Nelson et al. (2019); *Staphylococcus* with 0.01% relative abundance of identified genera in blow samples and *Streptococcus* with 0.09%. Identification of taxa with pathogenic potential may, therefore, be of special relevance in assessing the health status of cetacean populations, such as the SFPW targeted in this case study. Here, the identification of various pathogenic-potential genera across all samples may suggest that this resident population of SFPW might be vulnerable to respiratory disease. Nevertheless, the low number of samples and low definition in the microbiome results (not being able to identify the pathogenic species as previously discussed) do not allow us to properly infer the population vulnerability. We would also require additional data from transient individuals found within the study area and/or additional samples from the same species, but in other areas to eliminate the hypothesis that this microbiome might be a baseline within healthy individuals of this species. Other health assessment methodologies, such as nucleic acid-derived indices, have also recently been tested to study the ecophysiological traits of these animals that commonly occur in the surroundings of the Island of Madeira (Alves et al. 2020). In this cited study, the authors concluded that this SFPW population showed good ecophysiological conditions, although significantly lower when compared to other species from the same study area, which could be due to interspecific variations and not environmental conditions. Thus, the present blow microbiome work then provides complementary information to evaluate the health status of this marine mammal population.

Regarding the comparison of the core microbiome between the different datasets used, our results show that V4-V5 provides less variation in the data obtained from all the samples, with all the blow microbiomes showing similarity between them. Furthermore, it is clear that, within the same differentiated cluster for this dataset, the NMDS values for the sampled individuals which travelled together (Gma_2/Gma_3, Gma_4/Gma_5 and Gma_7/Gma_P1/Gma_P2) are tendentiously close. This reinforces the influence of sociality in the microbiome composition, similar to what was inferred in the beta-diversity analysis. Therefore, our results suggest that the V4-V5 dataset could be more consistent in determining the core microbiome present in the respiratory tract of free-ranging SFPW, showing less variation in this parameter of tested

samples. Nevertheless, the higher variation documented by V3-V4 may also be relevant since it provides a broader level of representation of the detected taxa within the blow. In this regard, a combination between the two primer sets could be complementary and represent a more robust way of characterising the microbiome. This is a widely used approach for metabarcoding studies as the combination of different primers have shown to improve taxa coverage, while also helping to reduce diversity bias (Lee et al. 2023).

Conclusions

The comparison between the primer sets showed that all samples from both datasets (V3-V4 and V4-V5) shared the main taxa composed of Actinobacteria, Bacteroidetes, Firmicutes and Proteobacteria phyla. This study provides a detailed characterisation of the microbial richness present in the blow of SFPW across multiple taxonomic levels. Following this work's results, it is concluded that the primer set selection for the microbiome assessment in cetacean blow samples should depend mainly on the goal of the analyses. If the main goal is to capture more diversity present in higher relative abundance, the V3-V4 primer set is demonstrated to have a better performance; whereas, if the purpose is to gather more information in the form of unique ASVs and to identify the microbial rare biosphere, we propose the use of the primer set targeting the hypervariable regions V4-V5. Despite the V4-V5 dataset detecting a higher number of unique ASVs and taxa, most had a relative abundance of < 1%. The V4-V5 dataset showed more consistent results in determining the core microbiome in the blow samples, while V3-V4 had higher variation. In this regard, a combination of primers may prove to be the more robust way to characterise the blow microbiome. This study also offers evidence that social behaviour influences the microbiome composition of the respiratory tract in cetacean species.

Nevertheless, several other aspects require consideration and future development to advance the blow microbiome as a health monitoring tool for cetaceans. Besides optimisation of the sampling and processing protocols, it is also relevant to test different methodologies to enhance sequencing efficiency and downstream procedures. Moreover, crossing this type of data with photogrammetry datasets to assess body condition is relevant to properly infer the pathogenic potential of these microbial communities in cetacean species.

In conclusion, this study represents an important contribution towards understanding the microbiome present in the respiratory tract of free-ranging cetaceans and marks the initial step in characterising the blow microbiome of SPFW. Finally, this study further underscores the potential of the blow microbiome as a future biomarker for assessing the health status and physiological state of the airways in free-ranging cetaceans.

Additional information

Conflict of interest

The authors have declared that no competing interests exist.

Ethical statement

No ethical statement was reported.

Funding

Author contributions

Conceptualization: SM, ÁG, FA, AMC, MR, AD, RV, CC, MPT, RF, CM. Data curation: BS, MPT. Formal analysis: MPT, BS. Funding acquisition: MR, RF, AMC, CM, ISP, LFCC, FA, AD. Investigation: MPT, ÁG, RV, BS, LA, AMC. Methodology: AMC, BS, FA, AD, RF, RV, ÁG, MR, MPT. Resources: FA, AD, MPT, CM, MR, ISP, RF, LFCC. Supervision: MPT, AMC, ISP, LFCC, CM. Visualization: MPT, BS. Writing - original draft: BS, LA. Writing - review and editing: SM, CM, BS, MR, ISP, MPT, AMC, FA, LFCC, ÁG, RF, RV, CC, AD, LA.

Author ORCIDs

Luís Afonso ⓘ https://orcid.org/0000-0003-3978-1388
Filipe Alves ⓘ https://orcid.org/0000-0003-3752-2745
Ana Dinis ⓘ https://orcid.org/0000-0002-1508-9197
Rita Ferreira ⓘ https://orcid.org/0000-0002-1383-9054
Ana M. Correia ⓘ https://orcid.org/0000-0003-4781-6894
Raul Valente ⓘ https://orcid.org/0000-0001-8544-373X
Ágatha Gil ⓘ https://orcid.org/0000-0003-4589-2201
Luis Filipe C. Castro ⓘ https://orcid.org/0000-0001-7697-386X
Isabel Sousa-Pinto ⓘ https://orcid.org/0000-0002-9231-0553
Massimiliano Rosso ⓘ https://orcid.org/0000-0002-5161-6767
Sandro Mazzariol ⓘ https://orcid.org/0000-0002-4756-1871
Maria Paola Tomasino ⓘ https://orcid.org/0000-0003-3255-0982

Data availability

The raw sequencing data from this study are openly available in the European Nucleotide Archive (ENA) with the reference number PRJEB72700 (https://www.ebi.ac.uk/ena/browser/home.

References

Acevedo-Whitehouse K, Rocha-Gosselin A, Gendron D (2010) A novel non-invasive tool for disease surveillance of free-ranging whales and its relevance to conservation programs. Animal Conservation 13(2): 217–225. https://doi.org/10.1111/j.1469-1795.2009.00326.x

Aguilar de Soto N, Alves F (2023) Short-Finned Pilot Whale *Globicephala macrorhynchus* Gray, 1846. Handbook of the Mammals of Europe. Springer International Publishing, Cham, 1–32. https://doi.org/10.1007/978-3-319-65038-8_101-1

Alves F, Quérouil S, Dinis A, Nicolau C, Ribeiro C, Freitas L, Kaufmann M, Fortuna C (2013) Population structure of short-finned pilot whales in the oceanic archipelago of Madeira based on photo-identification and genetic analyses: Implications for conservation. Aquatic Conservation 23(5): 758–776. https://doi.org/10.1002/aqc.2332

Alves F, Alessandrini A, Servidio A, Mendonça AS, Hartman KL, Prieto R, Berrow S, Magalhães S, Steiner L, Santos R, Ferreira R, Pérez JM, Ritter F, Dinis A, Martin V, Silva M, Aguilar de Soto N (2019) Complex biogeographical patterns support an ecological connectivity network of a large marine predator in the north-east Atlantic. Diversity & Distributions 25(2): 269–284. https://doi.org/10.1111/ddi.12848

Alves F, Dromby M, Baptista V, Ferreira R, Correia AM, Weyn M, Valente R, Froufe E, Rosso M, Sousa-Pinto I, Dinis A, Dias E, Teodósio MA (2020) Ecophysiological traits of highly mobile large marine predators inferred from nucleic acid derived indices. Scientific Reports 10(1): 4752. https://doi.org/10.1038/s41598-020-61769-7

Apprill A, Miller CA, Moore MJ, Durban JW, Fearnbach H, Barrett-Lennard LG (2017) Extensive Core Microbiome in Drone-Captured Whale Blow Supports a Framework for Health Monitoring. mSystems 2(5): e00119–e17. https://doi.org/10.1128/mSystems.00119-17

Atkinson S, Rogan A, Baker CS, Dagdag R, Redlinger M, Polinski J, Urban J, Sremba A, Branson M, Mashburn K, Pallin L, Klink A, Steel D, Bortz E, Kerr I (2021) Genetic, endocrine, and microbiological assessments of blue, humpback and killer whale health using unoccupied aerial systems. Wildlife Society Bulletin 45(4): 654–669. https://doi.org/10.1002/wsb.1240

Ballance LT (2018) Cetacean ecology. Encyclopedia of marine mammals (3rd edn). Academic Press, 172–180. https://doi.org/10.1016/B978-0-12-804327-1.00087-X

Betty EL, Zwamborn EMJ, Weyn M, Luck E, Alves F (2023) Life History Parameters, Sociobiology, and Reproductive Strategies of Pilot Whales. In: Würsig B, Orbach DN (Eds) Sex in Cetaceans. Springer, Cham, 327–351. https://doi.org/10.1007/978-3-031-35651-3_15

Bik EM, Costello EK, Switzer AD, Callahan BJ, Holmes SP, Wells RS, Carlin KP, Jensen ED, Venn-Watson S, Relman DA (2016) Marine mammals harbor unique microbiotas shaped by and yet distinct from the sea. Nature Communications 7(1): 1–13. https://doi.org/10.1038/ncomms10516

Bogomolni AL, Gast RJ, Ellis JC, Dennett M, Pugliares KR, Lentell BJ, Moore MJ (2008) Victims or vectors: A survey of marine vertebrate zoonoses from coastal waters of the Northwest Atlantic. Diseases of Aquatic Organisms 81(1): 13–38. https://doi.org/10.3354/dao01936

Boran J, Heimlich S (2019) Pilot whales: delphinid matriarchies in deep seas. Ethology and Behavioral Ecology of Odontocetes, 281–304. https://doi.org/10.1007/978-3-030-16663-2_13

Bray JR, Curtis JT (1957) An ordination of the upland forest communities of southern Wisconsin. Ecological Monographs 27(4): 326–349. https://doi.org/10.2307/1942268

Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP (2016) DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods 13(7): 581–583. https://doi.org/10.1038/nmeth.3869

Callahan BJ, McMurdie PJ, Holmes SP (2017) Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. The ISME Journal 11(12): 2639–2643. https://doi.org/10.1038/ismej.2017.119

Centelleghe C, Carraro L, Gonzalvo J, Rosso M, Esposti E, Gili C, Bonato M, Pedrotti D, Cardazzo B, Povinelli M, Mazzariol S (2020) The use of Unmanned Aerial Vehicles

(UAVs) to sample the blow microbiome of small cetaceans. PLoS ONE 15(7): e0235537. https://doi.org/10.1371/journal.pone.0235537

Colpitts J, McLoughlin PD, Poissant J (2022) Runs of homozygosity in Sable Island feral horses reveal the genomic consequences of inbreeding and divergence from domestic breeds. BMC Genomics 23(1): 1–17. https://doi.org/10.1186/s12864-022-08729-9

Cuvertoret-Sanz M, López-Figueroa C, Byrne AO, Canturri A, Martí-Garcia B, Pintado E, Pérez L, Ganges L, Cobos A, Abarca ML, Raga JA, Van Bressem M, Domingo M (2020) Causes of cetacean stranding and death on the Catalonian coast (western Mediterranean Sea), 2012–2019. Diseases of Aquatic Organisms 142: 239–253. https://doi.org/10.3354/dao03550

Díaz-Delgado J, Fernández A, Sierra E, Sacchini S, Andrada M, Vela AI, Quesada-Canales Ó, Paz Y, Zucca D, Groch K, Arbelo M (2018) Pathologic findings and causes of death of stranded cetaceans in the Canary Islands (2006–2012). PLoS ONE 13(10): e0204444. https://doi.org/10.1371/journal.pone.0204444

Esteban R, Verborgh P, Freitas L (2022) Dynamics of short-finned pilot whales long-term social structure in Madeira. Mammalian Biology 102(4): 1315–1332. https://doi.org/10.1007/s42991-022-00280-0

Evans P (2018) Habitat pressures. Encyclopedia of marine mammals (3rd edn). Academic Press, 441–446. https://doi.org/10.1016/B978-0-12-804327-1.00008-X

Foley C, O'Farrelly C, Meade KG (2011) Comparative analyses of the quality and yield of genomic DNA from invasive and noninvasive, automated and manual extraction methods. Journal of Dairy Science 94(6): 3159–3165. https://doi.org/10.3168/jds.2010-3987

Geoghegan JL, Pirotta V, Harvey E, Smith A, Buchmann JP, Ostrowski M, Eden J, Harcourt R, Holmes EC (2018) Virological sampling of inaccessible wildlife with drones. Viruses 10(6): 300. https://doi.org/10.3390/v10060300

Higgins R (2000) Bacteria and fungi of marine mammals: A review. The Canadian Veterinary Journal. La Revue Veterinaire Canadienne 41(2): 105.

Hunt KE, Moore MJ, Rolland RM, Kellar NM, Hall AJ, Kershaw J, Raverty SA, Davis CE, Yeates LC, Fauquier DA, Rowles TK, Kraus SD (2013) Overcoming the challenges of studying conservation physiology in large whales: A review of available methods. Conservation Physiology 1(1): cot006. https://doi.org/10.1093/conphys/cot006

Illumina (2013) 16S Metagenomic Sequencing Library Preparation Preparing 16S Ribosomal RNA Gene Amplicons for the Illumina MiSeq System. Illumina Technical Document, (Part. No. 15044223 Rev. B.).

Johnson WR, Torralba M, Fair PA, Bossart GD, Nelson KE, Morris PJ (2009) Novel diversity of bacterial communities associated with bottlenose dolphin upper respiratory tracts. Environmental Microbiology Reports 1(6): 555–562. https://doi.org/10.1111/j.1758-2229.2009.00080.x

Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Glöckner FO (2013) Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Research 41(1): 1–11. https://doi.org/10.1093/nar/gks808

Kumpitsch C, Koskinen K, Schöpf V, Moissl-Eichinger C (2019) The microbiome of the upper respiratory tract in health and disease. BMC Biology 17(1): 1–20. https://doi.org/10.1186/s12915-019-0703-z

Lahti L, Shetty S (2017) microbiome R package. Bioconductor. https://doi.org/10.18129/B9.bioc.microbiome

Lee HB, Jeong DH, Cho BC, Park JS (2023) Comparative analyses of eight primer sets commonly used to target the bacterial 16S rRNA gene for marine metabarcoding-based studies. Frontiers in Marine Science 10: 1199116. https://doi.org/10.3389/fmars.2023.1199116

Lima N, Rogers T, Acevedo-Whitehouse K, Brown MV (2012) Temporal stability and species specificity in bacteria associated with the bottlenose dolphins respiratory system. Environmental Microbiology Reports 4(1): 89–96. https://doi.org/10.1111/j.1758-2229.2011.00306.x

Lombardo MP (2008) Access to mutualistic endosymbiotic microbes: An underappreciated benefit of group living. Behavioral Ecology and Sociobiology 62(4): 479–497. https://doi.org/10.1007/s00265-007-0428-9

McMurdie PJ, Holmes S (2013) phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. PLoS ONE 8(4): e61217. https://doi.org/10.1371/journal.pone.0061217

McNichol J, Berube PM, Biller SJ, Fuhrman JA (2021) Evaluating and Improving SSU rRNA PCR Primer Coverage for Bacteria, Archaea, and Eukaryotes Using Metagenomes from Global Ocean Surveys. mSystems 3(6): e0056521. https://doi.org/10.1128/mSystems.00565-21

Mello DMD, Oliveira CA (2016) Biological matrices for sampling free-ranging cetaceans and the implications of their use for reproductive endocrine monitoring. Mammal Review 46(2): 77–91. https://doi.org/10.1111/mam.12055

Neary MT, Neary JM, Lund GK, Garry FB, Holt TN, Mohun TJ, Breckenridge RA (2014) A comparison of DNA collection methods in cattle and yaks. Journal of Animal Science 92(9): 3811–3815. https://doi.org/10.2527/jas.2013-7445

Nelson TM, Wallen MM, Bunce M, Oskam CL, Lima N, Clayton L, Mann J (2019) Detecting respiratory bacterial communities of wild dolphins: Implications for animal health. Marine Ecology Progress Series 622: 203–217. https://doi.org/10.3354/meps13055

Nicol C, Bejder L, Green L, Johnson C, Keeling L, Noren D, Van der Hoop J, Simmonds M (2020) Anthropogenic threats to wild cetacean welfare and a tool to inform policy in this area. Frontiers in Veterinary Science 7: 57. https://doi.org/10.3389/fvets.2020.00057

O'Mahony ÉN, Sremba AL, Keen EM, Robinson N, Dundas A, Steel D, Wray J, Baker CS, Gaggiotti OE (2024) Collecting baleen whale blow samples by drone: A minimally intrusive tool for conservation genetics. Molecular Ecology Resources: 13957. https://doi.org/10.1111/1755-0998.13957

Olson PA (2009) Pilot Whales: *Globicephala melas* and *G. macrorhynchus*. Encyclopedia of marine mammals (2nd edn). Academic Press, 847–852. https://doi.org/10.1016/B978-0-12-373553-9.00197-8

Parada AE, Needham DM, Fuhrman JA (2016) Every base matters: Assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. Environmental Microbiology 18(5): 1403–1414. https://doi.org/10.1111/1462-2920.13023

Parsons ECM, Baulch S, Bechshoft T, Bellazzi G, Bouchet P, Cosentino AM, Godard-Codding CAJ, Hoffmann-Kuhnt M, Hoyt E, Livermore S, MacLeod CD, Matrai E, Munger L, Ochiai M, Peyman A, Recalde-Saldas A, Regnery R, Rojas-Bracho L, Salgado-Kent CP, Slooten E, Wang JY, Wilson SC, Wright AJ, Young S, Zwamborn E, Sutherland WJ (2015) Key research questions of global importance for cetacean conservation. Endangered Species Research 27(2): 113–118. https://doi.org/10.3354/esr00655

Pirotta V, Smith A, Ostrowski M, Russell D, Jonsen ID, Grech A, Harcourt R (2017) An economical custom-built drone for assessing whale health. Frontiers in Marine Science 4: 425. https://doi.org/10.3389/fmars.2017.00425

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO (2013) The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. Nucleic Acids Research 41(D1): 590–596. https://doi.org/10.1093/nar/gks1219

Ribeiro H, de Sousa T, Santos JP, Sousa AGG, Teixeira C, Monteiro MR, Salgado P, Mucha AP, Almeida CMR, Torgo L, Magalhães C (2018) Potential of dissimilatory nitrate reduction pathways in polycyclic aromatic hydrocarbon degradation. Chemosphere 199: 54–67. https://doi.org/10.1016/j.chemosphere.2018.01.171

Robinson CV, Nuuttila HK (2020) Don't hold your breath: Limited DNA capture using non-invasive blow sampling for small cetaceans. Aquatic Mammals 46(1): 32–41. https://doi.org/10.1578/AM.46.1.2020.32

Robles-Malagamba MJ, Walsh MT, Ahasan MS, Thompson P, Wells RS, Jobin C, Fodor AA, Winglee K, Waltzek TB (2020) Characterization of the bacterial microbiome among free-ranging bottlenose dolphins (*Tursiops truncatus*). Heliyon 6(6): e03944. https://doi.org/10.1016/j.heliyon.2020.e03944

Sacco J, Ruplin A, Skonieczny P, Ohman M (2017) Polymorphisms in the canine monoamine oxidase a (MAOA) gene: Identification and variation among five broad dog breed groups. Canine Genetics and Epidemiology 4(1): 1–8. https://doi.org/10.1186/s40575-016-0040-2

Takahashi S, Tomita J, Nishioka K, Hisada T, Nishijima M (2014) Development of a prokaryotic universal primer for simultaneous analysis of Bacteria and Archaea using next-generation sequencing. PLoS ONE 9(8): e105592. https://doi.org/10.1371/journal.pone.0105592

Vendl C, Ferrari BC, Thomas T, Slavich E, Zhang E, Nelson T, Rogers T (2019) Interannual comparison of core taxa and community composition of the blow microbiota from East Australian humpback whales. FEMS Microbiology Ecology 95(8): 102. https://doi.org/10.1093/femsec/fiz102

Vendl C, Slavich E, Nelson T, Acevedo-Whitehouse K, Montgomery K, Ferrari B, Thomas T, Rogers T (2020) Does sociality drive diversity and composition of airway microbiota in cetaceans? Environmental Microbiology Reports 12(3): 324–333. https://doi.org/10.1111/1758-2229.12835

Vendl C, Nelson T, Ferrari B, Thomas T, Rogers T (2021) Highly abundant core taxa in the blow within and across captive bottlenose dolphins provide evidence for a temporally stable airway microbiota. BMC Microbiology 21(1): 1–15. https://doi.org/10.1186/s12866-020-02076-z

Venn-Watson S, Smith CR, Jensen ED (2008) Primary bacterial pathogens in bottlenose dolphins *Tursiops truncatus*: Needles in haystacks of commensal and environmental microbes. Diseases of Aquatic Organisms 79(2): 87–93. https://doi.org/10.3354/dao01895

Venn-Watson S, Daniels R, Smith C (2012) Thirty year retrospective evaluation of pneumonia in a bottlenose dolphin *Tursiops truncatus* population. Diseases of Aquatic Organisms 99(3): 237–242. https://doi.org/10.3354/dao02471

Walters W, Hyde ER, Berg-lyons D, Ackermann G, Humphrey G, Parada A, Gilbert JA, Jansson JK (2015) Improved Bacterial 16S rRNA Gene (V4 and V4-5) and Fungal Internal Transcribed Spacer Marker Gene Primers for Microbial Community Surveys. mSystems 1(1): e0009–e0015.

Wickham H (2009) ggplot2: elegant graphics for data analysis. Springer-Verlag New York. https://doi.org/10.1007/978-0-387-98141-3

Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen T, Miller E, Bache S, Müller K, Ooms J, Robinson D, Seidel D, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019) Welcome to the Tidyverse. Journal of Open Source Software 4(43): 1686. https://doi.org/10.21105/joss.01686

Zehr JP, Weitz JS, Joint I (2017) How microbes survive in the open ocean. Science 357(6352): 646–647. https://doi.org/10.1126/science.aan5764

Supplementary material 1

Supplementary information

Authors: Beatriz Santos, Maria Paola Tomasino

Data type: pdf

Explanation note: **table S1**. Primers used for the blow microbiome analysis, targeting the 16S rRNA gene hypervariable region. **table S2**. Metadata of the blow samples collected and analysed: Sample ID, sampling date, geographical coordinates (DD), behaviour, age class, number of individuals sampled (1/pool) and residency pattern. Residency pattern was confirmed via Photo-ID, comparing with (OOM/MARE-ARDITI) Madeira's photographic-ID catalogue of the species following Alves et al. (2020): Resident - individuals that exhibited multiyear and year-round site fidelity, i.e. photographed in at least 4 years and all 4 seasons; Visitor - individuals that exhibited multi-year, but seasonal specific presence between July and December (i.e. covering two following seasons); Extended Visitor - individuals that exhibited multi-year but seasonal specific presence between July and March (i.e. covering three following seasons); Transient - individuals photographed once (i.e. not previously catalogued). Note: The residency pattern was attributed as a reference, but no statistical comparison was made due to the low number of individuals/samples in each category. **table S3**. Output of a Wilcoxon Signed-Rank Test, comparing the indexes of the short-finned pilot whale blow community richness obtained with two different primer sets (V3-V4 and V4-V5). **table S4**. Output of the PERMANOVA analysis (Df = degrees of freedom, SS = sum of squares, R^2 = coefficient of determination, PR (> F) = p-value) to test for the influence of the primer sets (V3-V4 and V4-V5) on the distribution of prokaryotic communities of the blows of short-finned pilot whales. **fig. S1**. Sampling kit used for blow sampling: A PERFORMAgeneTM PG-100 swab collection kit (DNA Genotek®); B Kit swab attached to an extendable 5-metre aluminium pole; C- Collection of blow samples of SFPW. **fig. S2.** Relative abundance of prokaryotic top 10 classes identified across the SFPW blow samples, by V3-V4 (on the top) and V4-V5 (on the bottom) datasets. **fig. S3**. Relative abundance of prokaryotic top 10 families identified across the SFPW blow samples, by V3-V4 (on the top) and V4-V5 (on the bottom) datasets. **fig. S4**. Major taxonomic families obtained for the short-finned pilot whale blow samples (in V3−V4 and V4−V5 datasets). The total sequence proportion of each family in the V3−V4 (left side) and V4−V5 (right side) datasets is represented in the x-axis. The numbers in each column represent the number of observed ASVs affiliated with each taxonomic family. Colour coding represents the different taxonomic classes. Only families that comprised at least 1% of sequences, in at least one of the datasets, were included in the figure.

Link: https://doi.org/10.3897/mbmg.8.121060.suppl1

Supplementary material 2

Upstream Bioinformatic Analysis

Authors: Beatriz Santos, Maria Paola Tomasino

Data type: R file

Explanation note: R script used for the upstream bioinformatic analysis (DADA2 software package) of Illumina MiSeq originated reads.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.8.121060.suppl2

Supplementary material 3

Downstream Bioinformatic Analysis

Authors: Beatriz Santos, Maria Paola Tomasino

Data type: R file

Explanation note: R script used for the downstream bioinformatic analysis (taxonomic identification, alpha and beta diversity) of quality-filtered Illumina MiSeq originated reads.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.8.121060.suppl3

Supplementary material 4

Bioinformatic Analysis: core microbiome

Authors: Beatriz Santos, Maria Paola Tomasino

Data type: R file

Explanation note: R script used for the calculation of the core microbiome for each sequenced dataset.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.8.121060.suppl4